How Opening Schools Influence COVID Infections – Empirical Evidence from Czech Republic

Martin Šmíd, Milan Zajíček*

May 15, 2021

Main findings

- Opening of schools in the Czech Republic strongly increases the infection numbers of students
- The effect of ordering masks in classrooms is significant in primary schools and in secondary schools
- Data do not support the hypothesis that the increase of reported cases after opening schools is due to more intense testing in schools
- Except for kindergartens, data do not support the hypothesis that, when being out of schools, children are infected elsewhere
- Simple indicator of safe opening schools can be constructed.
- We show that hypothetical opening of schools in the middle of the pandemics would not overturn the subsequent decreasing trend, yet it would increase the infections significantly.

Introduction

Closing schools during the present pandemics is the one of the most controversial issues. According to epidemiological mainstream, schools are strong drivers of respiratory diseases [cit]; therefore, closing schools was one of the first governmental reactions to the outburst of the present pandemic. However, the price we pay for closing schools is high. Online education is not an equivalent substitute for the in-person one; moreover, the isolation of students leaves deficit in their necessary social contacts [cit]. Thus, for the society, decision whether and

*Institute of Information Theory and Automation of the CAS, BISOP

when to close school means a painful trade-off, making any decision political rather then scientific; science, however, has irreplaceable role in giving the best possible (quantitative) basis for these decisions.

Unfortunately, it is difficult to evaluate the effect of closing schools during the present pandemic. The virus was completely unexplored as it came and its knowledge increased only gradually, so we only gradually get to know, whether and how differently the virus affects children in comparison with the rest of the population. To estimate the effect from running epidemic data is also problematic, mainly because the counter-epidemic measures are usually being introduced (and released) together, so it is difficult to distinguish their effects, and even in cases when the measures are applied in different times, still they can be assessed only in the context of the other measures applied. [Kulweit 2020], for instance, regards school closure as a very effective means of curbing the present pandemics; however, it is necessary to take into account that school closures were usually the first measures applied, so their measured effect may be overestimated.

In our analysis, we exploit a "gap" in this puzzle: the fact that certain age cohorts of children nearly uniquely correspond to the degrees of schools and, except for kindergartens, the majority of children do attend their classes. Thus, it could be expected that closing (opening) of particular degrees will result in significant changes in infections in the corresponding cohorts. The other measures, on the other hand, are usually much less age specific (wearing masks, bans of gatherings) or are affecting different age cohorts (home office) so they can be expected not to obscure the effects of school closures. Later we show that actual data prove us right in both these assumptions - the effects of closures are strongly statistically significant and the data do not exhibit co-linearity.

In the Czech Republic, all kindergartens, primary and secondary schools had been closed on March 12, 2020, shortly after the first cases of COVID infection appeared. While the kindergartens had been reopened without much restriction in the beginning of May, the primary and the secondary schools were opened during May only in a very limited regime until the summer vacation. After the vacation, all the schools had been opened for two weeks nearly without any restriction. Next, after the overall incidences started to rise, wearing masks in classrooms had been imposed. In the end of October, after serious increases of new cases, all the schools had been closed until April 2021, except for kindergartens, which remained open until March 2021, the first two grades of the elementary schools, which were opened from the half of November until February, and the final grades of the secondary schools, which were opened one month before Christmas. On April 12th, 2021, kindergartens opened for the final year and the other schools gradually started to be opened under rotation schemes. See the following Table for the schedule.

Date	Kindergartens	First degree	Second degree	Secondary
Mar-13	closed			
May-04	open	closed	closed	closed
May-11			last year	
May-25		max. 15 l.y.		max. 15 l.y.
Jun-01		max. 15 l.y.		
Jun-26	end of the school year			
Sep-01	start of the school year			
Oct-14	open	closed		
Nov-16	open	open 1^{st} 2^{nd}	closed	closed
Nov-23	open	open 1^{st} 2^{nd}	closed	last year
Mar-01	closed			

To evaluate the effects of the closures, we used a simple epidemiological-like model, in which the infections in the children cohorts depend on the overall state of epidemic plus the effect of schools, the former summand being linearly dependent on the product of the current magnitude of the epidemic and the current overall contact restriction, the latter being linearly dependent on the the product the current school restrictions and the number of infected in the cohort in question. We find the school closures as a whole as a significant inhibitor of the epidemic, with masks being provably efficient in the primary and secondary schools.

Next we construct an indicator sub-unit value of which guarantees keeping the infections in the cohort in question bounded or decreasing, if the bound is exceeded; this indicator can serve for decisions on school restriction degree during the epidemic.

Within the Discussion, we also discuss possible alternative explanation of the infection increases by more extensive testing in schools, and the hypothesis, that during school closures, children are increasingly infected elsewhere. We find, that, except for the latter hypothesis and kindergartens, data do not support these explanations.

Methods

In the Czech Republic, each September first, children who completed their sixth year are obliged to enter the first grade of an elementary school. When the parents wish, a child who is six as late as by the end of the year can enter the first class, too, some children, on the other hand, start their school later for various, mostly developmental, reasons. If we neglect these exceptions, we can conclude that the first year pupils are six or seven; consequently the first degree (first five grades) pupils belong to the age cohort $6-11$. The second degree (the sixth to the ninth year) students to the cohort $11-15$. Finally, the students of secondary schools, which typically take four year in the Czech Republic, belong to the cohort $15-19$. The lowest age for admission to a kindergarten is three, so the the vast majority of children attending kindergartens falls into the cohort $3-6$. For simplicity we assume that the frontier one-year cohorts split by half between the competing school categories, so the infections happening in the cohort split by two between each category. Thus the number of cases by pre-school children over week t will be $X_t^1 = Z_t^3 + Z_t^4 + Z_t^5 + \frac{1}{2} Z_t^6$, the number by the first degree $X_t^2 = \frac{1}{2} Z_t^6 + Z_t^7 + \cdots + \frac{1}{2} Z_t^{11}$, by the second degree $X_t^3 = \frac{1}{2} Z_t^{11} + Z_t^{12} + \cdots + \frac{1}{2} Z_t^{15}$ and that by the secondary schools $X_t^4 = \frac{1}{2} Z_t^{15} + Z_t^{16} + \cdots + \frac{1}{2} Z_t^{19}$,where Z_t^j is the number of cases by the j-year old individuals.

In addition to these four cohorts, we studied the infections in the first two grades of primary schools, which had been exclusively open during winter, taking $X_t^{2*} = \frac{1}{2} Z_t^6 + Z_t^7 + \frac{1}{2} Z_t^8$ as number of cases therein.

In line with the mainstream epidemiological modeling, we assume that the number of overall infections X_t is proportional to the previous number of infected and the overall contact reduction, where we take the last week incidence as a proxy for the former:

$$X_t \doteq r_t C_{t-1} X_{t-1};$$

here, C_{t-1} is the overall risk contact reduction within the overall population and r_t is the current growth rate given that no contact reduction takes place (including the effects of virus mutations, possible seasonal influences, natural immunization and vaccination).

For the school children we assume that their infection consists of a fraction of the overall infections and additional infections coming from schools. In particular, for the i-th school category,

$$X_t^i \doteq \alpha^i r_t C_t X_{t-1} + \gamma^i r_t S_{t-1}^i X_{t-1}^i, \qquad S_{t-1}^i = D_{t-1}^i (1 - \mu^i M_{t-1}^i), \tag{1}$$

where α^i and γ^i are constants, D_τ^i is the contact restriction at the school, M_τ^i is the indicator of wearing masks in classrooms and μ^i is the (unknown) efficiency of masks.

Further, assuming that, only a ratio c of cases is reported, we may divide (1) by c to get

$$Y_t \doteq r_t C_{t-1} Y_{t-1}, \qquad Y_t^i \doteq \alpha^i r_t C_{t-1} Y_{t-1} + \gamma^i r_t S_{t-1}^i Y_{t-1}^i, \qquad 1 \le i \le 4, \tag{2}$$

where $Y_t \doteq c X_t$ is the overall reported number of infections and $Y_t^i \doteq c X_t^i$ is the reported infections number in the i-th cohort. By dividing by the size of the cohort i, we get

$$P_t^i \doteq \beta^i r_t C_{t-1} P_{t-1} + \gamma^i r_t S_{t-1}^i P_{t-1}^i, \qquad P_t^i = \frac{Y_t^i}{s^i}, \qquad P_t = \frac{Y_t}{s}, \qquad \beta^i = \frac{s}{s^i}, \tag{3}$$

where s_i is the size of the i-th cohort and s is the whole population size.

Equations (2) and (3) may serve for understanding of infection spread in schools and consequently for policy recommendations. Assume that our goal is to keep new cases in the cohort less than some y_0 (for instance, corresponding to 50 cases per 100 thousand) and, if the number is higher, to decrease it by a ratio ρ_0. From (2) we get that, for it to happen at t, it has to be

$$\alpha^i r_t C_{t-1} Y_{t-1} + \gamma^i r_t S_{t-1}^i Y_{t-1}^i \le \max(\rho_0 Y_{t-1}^i, y_0)$$

which happens if

$$\rho_t := \rho_t^\alpha + \rho_t^\gamma < \rho_0, \qquad \rho_t^\alpha = \alpha^i r_t C_{t-1} \frac{Y_{t-1}}{\max(Y_{t-1}^i, y_0)}, \qquad \rho_t^\gamma = \gamma^i r_t S_{t-1}^i \min\left(1, \frac{Y_{t-1}^i}{y_0}\right)$$

The equivalent formulation (3), on the other hand, allows comparisons between cohorts, as it speaks in the language of infection probabilities: here the first/second summand evaluates the probability of infection outside/inside the school.

For the purpose of estimation, we may impose for S_{t-1}^i in (3) to get a linear model

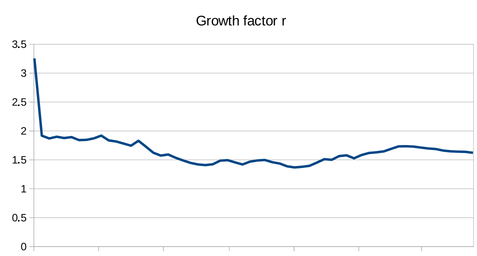
$$\begin{aligned} &P_t^i \doteq \beta^i Q_t + \gamma^i U_t + \delta^i V_t + \epsilon_t, \\ &Q_t = R_t P_t, \qquad U_t = r_t D_{t-1}^i P_{t-1}^i, \qquad V_t = r_t D_{t-1} M_{t-1} P_{t-1}, \qquad \delta^i = -\mu\gamma^i. \end{aligned} \tag{4}$$

Here, ϵ_t is centered error term for which we assume $\mathrm{var}(\epsilon_t) \sim Q_t^2$– see Appendix for further details and more rigorous treatment.

Data

Our data came from public sources. The series values $Y^1, \ldots, Y^4, Y^{2*}$ and Y are taken from the public dataset by the Czech Ministry of Health, namely the anonymized person-level list of reported infections (cit osoby.htm) including, among other things, the date of reporting the infection and the age.

The growth rate r_t is computed by exponential smoothing of series $\frac{Y_t/Y_{t-1}}{C_{t-1}}$ with parameter $\alpha = 0.1$, as shown in the following chart.



The values of S and M are estimated from publicly available sources, mostly from resolutions of the Czech government concerning school attendance, which have been put into effect through decrees of the Ministry of Education (see the electronic supplement TBD for details). Only weeks starting from April 6th, 2020 are taken into account, as we regard the previous epidemic data as noisy, unreliable and suffering from small sample properties. The last values correspond to the week staring on April 5th, 2021. The values of S were set to 1 if the school was open without any restrictions other than mask, and to 0 if they were closed. In the other weeks, we estimated values of S only when we could determine S objectively. Thus, we excluded observations corresponding to holidays, weeks with no more than two possible school days, and the weeks with non-standard regime, such as rotations or limited numbers of students in a room. We also excluded the weeks where only first two classes had been open, which we study separately.

The corresponding values of S and M are listed in the Table in Appendix together with the notes on how we determined fractional values S and reasons of observation exclusions. The following table summarizes our data-set.

	Kindergartens	First degree	Second degree	Secondary grades	First two
Observaions	42	23	34	35	26
Opened	32	7	7	9	12
... with masks	0	5	5	7	12
... without masks	32	2	2	2	0

Results

For all the cohorts with observations both with and without masks, which are both the primary ones and the secondary one, we estimated β,γ and δ in (4). In all three cases, the coefficients came out "reasonably": β came out undoubtedly significant (overall epidemics influences the infections), $\gamma > 0$ (schools add the infections) and $0 \leq -\delta \leq \gamma$ (masks reduce infection in schools). For the kindergartens, where masks were never worn in classes, and for the first two classes, where masks were always worn, we subsequently estimated (4) with $\delta^1 = 0$. In both the cases, the coefficient γ (the influence of opening) came out undoubtedly significant.

The results of estimation may be found in the following Table.

	Kindergartens	First degree	Second degree	Secondary	1. and 2.
β	$0.369^{***}(0.0578)$	$0.55^{***}(0.0405)$	$0.83^{***}(0.0397)$	$0.57^{***}(0.048)$	$0.503^{***}(0.0419)$
γ	$0.41^{***}(0.055)$	$0.61^{**}(0.262)$	$0.62^{***}(0.208)$	$1.11^{***}(0.154)$	$0.36^{***}(0.042)$
δ		$-0.49^{*}(0.251)$	$-0.6^{**}(0.222)$	$-0.85^{***}(0.177)$	

The results strongly indicate the influence of opening corresponding school categories and they also indicate that wearing masks reduces infections significantly. Unfortunately, due to small numbers of observations, the estimates of γ and δ are rather imprecise. The values of the variance inflation factors below 5 in each case, however, suggest that the models do not suffer from colinearity, so they should be capable to distinguish sharply the effect of wearing masks.

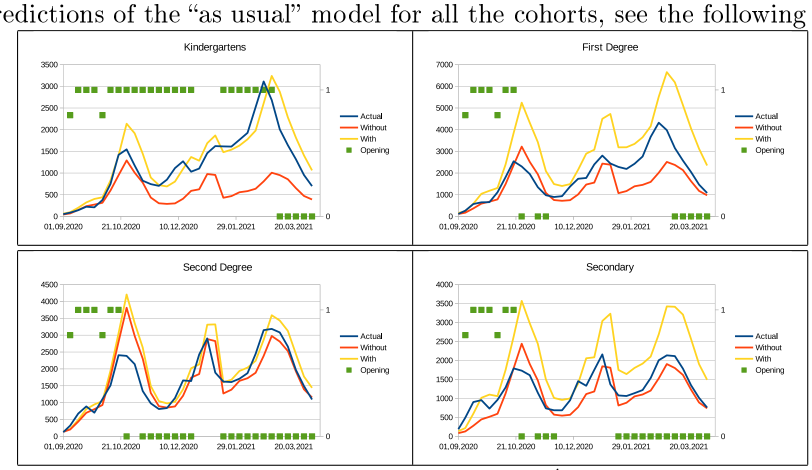
When comparing the results of the individual cohorts, the values of β, indicating the rate of infection outside schools, increase with the age, surprisingly dropping at the secondary schools. The values γ (the influence of school opening without masks) increase with age. The estimated effect of wearing masks (coefficients δ) increases with age.

Further we focused focused on the prevailing modus operandi of the individual school categories, namely not wearing masks in kindergartens and wearing masks elsewhere. To this end, we re-estimated (4) with $\delta = 0$ using only records where either schools were closed or the usual modus operandi (i.e. without masks in kindergartens, with masks otherwise) took place. The results are in the following table

	Kindergartens no masks	First degree masks	Second degree masks	Secondary masks	Primary masks
β	$0.369^{***}(0.0578)$	$0.55^{***}(0.0429)$	$0.829^{***}(0.0419)$	$0.57^{***}(0.0425)$	$0.503^{***}(0.0419)$
γ	$0.41^{***}(0.055)$	$0.19(0.133)$	$0.12(0.103)$	$0.44^{***}(0.08)$	$0.36^{***}(0.039)$
α	$0.012^{***}(0.0018)$	$0.029^{***}(0.0023)$	$0.035^{***}(0.0017)$	$0.022^{***}(0.0016)$	$0.011^{***}(0.0009)$

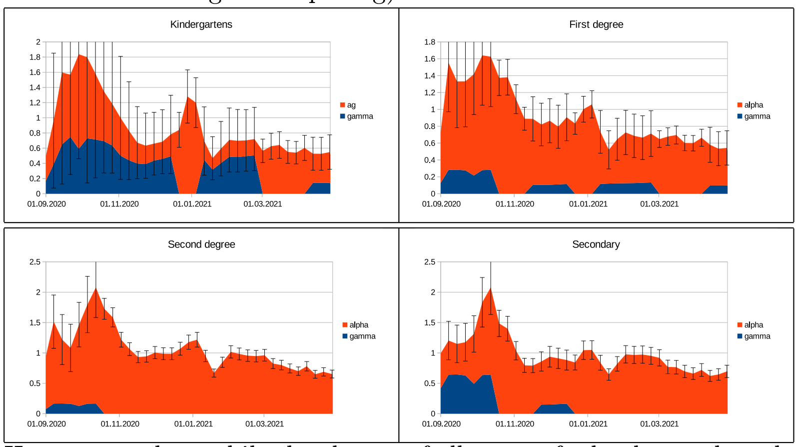
The values and errors of β are identical for the kindergartens and the first grades (the same observations were used) and are nearly identical in the remaining cases. The values of γ are consistent with the previous results, roughly corresponding to $\gamma - \delta$ of the previous model (in the previous model, γ measures the influence of not wearing masks, while here it concerns to usual modus operandi). For the first and the second degrees, the values are insignificant, which is non surprising, as, by the previous model, their "true" values are evidently too close to zero to be significant with the present limited dataset. The (insignificant) result at the second degree indicates less contribution than the first degree. In the secondary schools, the contribution is again higher.

To give the reader an idea about quantities in question, we further plotted predictions of the "as usual" model for all the cohorts, see the following Figure

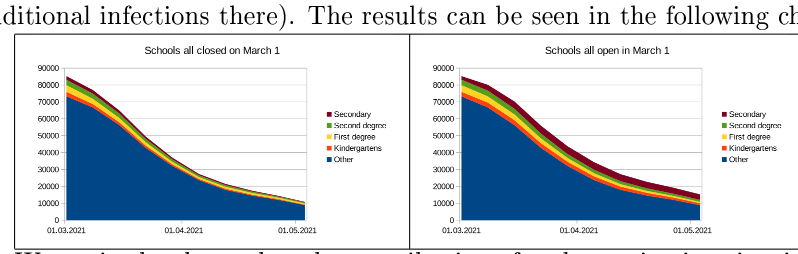


The graphs show one-week predictions of Y_t^i with– and without schools (fully) open, i.e. with $S_{t-1} = 1$, $S_{t-1} = 0$, respectively, and with M set according to the usual manner. In addition, actual (observed) numbers and the degree of opening schools one week before (values of S_{t-1}^i) are depicted. It can be seen that, according to the model, full opening of schools (in a usual manner) at least doubles the new cases in corresponding except for the second degrees; however, it is important to keep in mind the insignificance of results for the primary schools.

Finally, we evaluated index ρ_t for all the types during the school year 2020/21. Recall that ρ_t is an indicator of safe opening of schools, whose value below 1 indicate that the cases in the corresponding cohort will not rise. Recall also that the indicator is a sum of a part dependent on the outside epidemic (indexed by α) and the part dependent on the degree of opening of schools (indexed by γ). The following graphs show the values of the index through time (with the the true degree of opening)



Here we see that, while the closure of all types of schools may have helped in Autumn, their closure, or not opening at least in some restricted mode, in March seems strict. However, it should be stressed that the indicators ρ_t speaks only about the growth at a given time; if the schools were open, however, the numbers of infected students would rise, possibly amplifying the effect. To examine this, we modeled the hypothetical situation in which all the schools were all open (in a usual manner, i.e. with masks wherever except for kindergartens), on March 1, which was the time when all the schools had been closed in reality. In our forecast, we assumed the numbers of cases outside the examined cohorts to be as in reality while the numbers in the cohorts were computed according to the "as usual" model (i.e. we neglect a possible increase outside schools due to additional infections there). The results can be seen in the following charts.



We see in the chart, that the contribution of such opening is quite significant and, yet it does not seem to overturn the decreasing trend of the epidemic, it probably would, if other areas acted according the same logic and opened too. The result however shows that, if opening schools were politically prioritized, it would not make much harm.

Discussion

In our honest opinion, our results convincingly prove the strong influence of in-class education over various types of schools on the epidemic spread, as well as its reduction by wearing masks in rooms. However, our study has limitations. First of all it is the shortage of data. Most of the weeks we examine the schools have been closed, and when they were opened, mostly with masks, which clearly complicates quantification of their influences. Therefore, the quantitative results should be taken with slight caution.

Further, there is a determinant of children infections not taken into account in our analysis: encounters with teachers, which can be significant (TBD cite Neruda et al). This influence could be taken into account by adding another coefficient into (4); however, due to the data shortage, there would be little chance for significant results.

Further we discuss two arguments, often in discussions on schools: that children, not going to schools, are infected anyway in other environments, and that this influence of schools is spurious due to more extensive testing of students attending schools in person, (yet, in the examined period, no preventive testing in schools took place). In our model, the former hypothesis would mean that, in addition to fraction $\alpha^i Q_t$infected outside schools, $k^i(1-S^i_{t-1})Q_t$ for some k^i would be infected. The latter hypothesis, on the other hand, could be expressed as $Y^i_t = (c+S^i_{t-1}d^i)X^i_t$ for some d^i. Both the hypotheses can thus be examined by estimating a variant of the "as usual" model

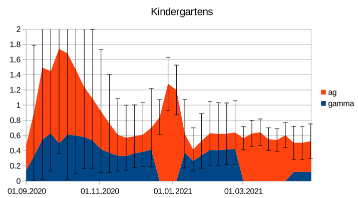
$$P^i_t \doteq \beta^i Q_t + \gamma^i U_t + \phi^i W^i_t + \eta_t, \qquad W^i_t = Q_t S^i_{t-1};$$

with values of $\phi^i < 0$ speaking for the first hypothesis (children are infected anyway), the values $\phi > 0$ for the second one (there is higher chance to be tested at school). The results of this estimation can be found in the following Table

	Kindergartens no masks	First degree masks	Second degree masks	Secondary masks	Primary masks
β	$0.535^{***}(0.075)$	$0.557^{***}(0.0385)$	$0.846^{***}(0.0454)$	$0.578^{***}(0.0484)$	$0.49^{***}(0.0466)$
γ	$0.49^{***}(0.058)$	$-0.74(0.467)$	$-0.17(0.339)$	$0.4^{*}(0.22)$	$0.35^{***}(0.047)$
ϕ	$-0.29^{***}(0.098)$	$0.64^{**}(0.322)$	$0.3(0.357)$	$0.04(0.305)$	$0.04(0.093)$

We can see from the table, that, for the two oldest cohorts, ϕ is insignificant, while, for the kindergarten cohort, results support the hypothesis "infected instead". For the first degree cohort, data suggest the "over-tested" hypothesis; here, however, the result is unintuitive ($\gamma < 0$), which is probably due colinearity (VIR is nearing 20), invalidating the estimate of ϕ, too. Thus, only "infected instead" alternative in kindergartens should be taken into account. However, the results are comparable here: the following charts show the "safety criterion" ρ_t, which is now changed to

$$\rho_t := \rho^{\alpha\psi}_t + \rho^{\gamma}_t < \rho_0, \qquad \rho^{\alpha\psi}_t = r_t(\alpha^i C_{t-1} + \psi^i S^i_{t-1}) \frac{Y_{t-1}}{\max(Y^i_{t-1}, y_0)}, \qquad \psi^{i} = \frac{s^i}{s}\phi^i$$



Questions/Ideas

- Not split primary schools (and present degrees only as secondary results)?

Appendix

The Statistical Model

TBD

Date	Kindergartens		First degree		Second degree		Secondary		Primary	
	S^1	M^1	S^2	M^2	S^3	M^3	S^4	M^4	S^{2*}	M^{2*}
06-Apr-20	0		0		0		0		0	
13-Apr-20	0		0		0		0		0	
20-Apr-20	0		0		0		0		0	
27-Apr-20	0		0		0		0		0	
04-May-20	1	N	0		0		0		0	
11-May-20	1	N	0		$\times^{ad}$		0.25^d	Y	$\times$	
18-May-20	1	N	0		$\times^{ad}$		0.25^d	Y	$\times$	
25-May-20	1	N	$\times^a$		$\times^{ad}$		$\times^{ad}$		$\times$	
01-Jun-20	1	N	$\times^a$		$\times^{ad}$		$\times^{ad}$		$\times$	
08-Jun-20	1	N	$\times^a$		$\times^{ad}$		0		$\times$	
15-Jun-20	1	N	$\times^a$		$\times^{ad}$		0		$\times$	
22-Jun-20	1	N	$\times^a$		$\times^{ad}$		0		$\times$	
29-Jun-20	$\times^e$		$\times^e$		$\times^e$		$\times^e$		$\times$	
	$\vdots$		$\vdots$		$\vdots$		$\vdots$		$\times$	
24-Aug-20	$\times^e$		$\times^e$		$\times^e$		$\times^e$		$\times$	
31-Aug-20	0.8^a	N	0.8^b	N	0.8^b	N	0.8^b	N	$\times$	
07-Sep-20	1	N	1	N	1	N	1	N	$\times$	
14-Sep-20	1	N	1	Y	1	Y	1	Y	$\times$	
21-Sep-20	1	N	1	Y	1	Y	1	Y	$\times$	
28-Sep-20	0.8^a	N	0.8^b	Y	0.8^b	Y	0.8^b	Y	$\times$	
05-Oct-20	1	N	1	Y	1	Y	1	Y	$\times$	
12-Oct-20	1	N	1	Y	1	Y	1	Y	$\times$	
19-Oct-20	1	N	0		0		0		0	
26-Oct-20	1	N	$\times^g$		$\times^g$		$\times^g$		$\times$	
02-Nov-20	1	N	0		0		0		0	
09-Nov-20	1	N	0		0		0		0	
16-Nov-20	1	N	$\times^c$		0		0		1	Y
23-Nov-20	1	N	$\times^c$		0		$\times^d$		1	Y
30-Nov-20	1	N	$\times^c$		0		$\times^d$		1	Y
07-Dec-20	1	N	$\times^c$		0		$\times^d$		1	Y
14-Dec-20	1	N	$\times^c$		0		$\times^d$		1	Y
21-Dec-20	$\times^f$		$\times^f$		$\times^f$		$\times^f$		$\times$	
28-Dec-20	$\times^f$		$\times^f$		$\times^f$		$\times^f$		$\times$	
04-Jan-21	$\times^f$		$\times^f$		$\times^f$		$\times^f$		$\times$	
11-Jan-21	1	N	$\times^c$		0		0		1	Y
18-Jan-21	1	N	$\times^c$		0		0		1	Y
25-Jan-21	1	N	$\times^c$		0		0		1	Y
01-Feb-21	1	N	$\times^c$		0		0		1	Y
08-Feb-21	1	N	$\times^c$		0		0		1	Y
15-Feb-21	1	N	$\times^c$		0		0		1	Y
22-Feb-21	1	N	$\times^c$		0		0		1	Y
01-Mar-21	0		0		0		0		0	
08-Mar-21	0		0		0		0		0	
15-Mar-21	0		0		0		0		0	
22-Mar-21	0		0		0		0		0	
29-Mar-21	0		0		0		0		0	
05-Apr-21	0		0		0		0		0	

Table 1: Notes: a – only 15 pupils in the classroom (approx. half), b – only 4 days from week, c – only 1st and 2nd calsses are open, d – only the last year open, e – summer vacation, f – Christmas vacation, g – autumn vacation